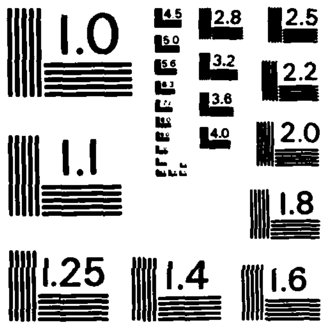
AD-A161 964 INITIALIZATION EFFECTS IN COMPUTER SIMULATION EXPERIMENTS(U) CORNELL UNIV ITHACA N Y COLL OF ENGINEERING L SCHRUBEN ET AL. SEP 84 TR-594 1/1

UNCLASSIFIED N00014-81-K-0037 F/G 9/2 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

AD-A161 964

SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 594

October, 1983
(Revised September, 1984)

INITIALIZATION EFFECTS IN COMPUTER SIMULATION EXPERIMENTS

by

Lee Schruben

and

David Goldsman

This research was supported by the Office of Naval Research under contract N00014-81-k-0037. David Goldsman is currently in the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332.

Abstract

Issues relating to initializing a simulation program are discussed, and suggested procedures for handling this source of error are reviewed.

1. Introduction

When a simulation program is run on a computer, initial values for all variables must be specified. The experimenter typically does not know what initial values are appropriate for all of the variables in the program. These values are often selected in a rather arbitrary manner as a matter of convenience; such a selection might have a significant influence on the outcome of the experiment. If the setting of the initial conditions for a run has a major but unrecognized effect, then the results from the run can be misleading. Thus, initialization can be a serious source of error in simulation experiments.

The simulation literature contains many papers on *'the initialization bias problem'*. Probably as much has been written about this topic as any other single area of simulation output analysis, but there does not seem to be any definitive solution to the problem. Indeed, no universally accepted definition of the problem exists. This is because there is not just a single issue in initialization of simulations; there are several.

One focus has been on obtaining accurate (low bias) and precise (low variability) estimators of some 'steady state' measure of simulated system performance. Say that the simulated process at time t, Y_t, has a distribution function denoted by F_t. The problem is to estimate properties (e.g., moments, quantiles) of the limiting (steady state) distribution, $F \equiv \lim_{t\to\infty} F_t$. The difficulty is that in many simulations (e.g., the M/M/1 queue with high traffic intensity), the limiting distribution is *asymptotic*. None of the data in the output series will be sampled from the distribution function F. This

problem has characteristics similar to forecasting problems [see Kimbler, *et al.* (1979) and Snell (1980)].

Another concern is that the effects of initialization do not lead to erroneous experimental conclusions. If the purpose of the simulation study is steady state system performance estimation, then the same steady state issues mentioned above may be pertinent. However, many simulation models are used as aids in (non-steady state) system design and evaluation. In design studies, the concern is whether a better design could be possible if initialization errors were not present. Evaluation studies question whether or not the presence of initialization errors favors other than the best of several alternatives that are being examined. In short, does the presence of initialization effects potentially result in an incorrect decision?

In Section 2 of this paper, some examples of simulation initialization effects are presented. Section 3 discusses the mathematical aspects of initializing a simulation program. Suggested procedures for detecting and handling this source of error are reviewed in Section 4.

2. Some examples

Some simple examples illustrate various problems that can arise when initializing a simulation program: If a simulation of a new factory is initiated with no work in progress, then the production of finished goods will be relatively low early in the run. Products have not yet had time to flow through the system (and so there will be little congestion in the system for some time). A naive statistical

analysis of the data from such a simulation might conclude that the production is positively correlated with congestion on the factory floor. This correlation is an artifact caused by the way in which the experiment was run; the correlation is not a natural characteristic of the system being studied. Nevertheless, the study might suggest that one way to increase production is to make the factory more congested. Low observed machine utilizations (perhaps again due to the simulated factory being initially empty) could lead the analyst to recommend that fewer machines than originally planned be installed. The use of fewer machines could then result in severe system bottlenecks as production increases.

Visually detecting initialization effects in estimators of system performance can sometimes be a non-trivial task. Figure 1 is the output (customer waiting times) of a simulated queueing system which was initially started with no customers in the system. The initial portion of the run does not appear to differ radically from the rest of the run. Actually, the average waiting time for the first 50 customers is 10 minutes while that for the first 500 customers is 27 minutes! The relatively low observed waiting times for early customers are due in part to the empty and idle initialization of the system.

We continue with the above queueing example. Suppose that the system closes each night, even if there are still customers waiting in line for service. If some of these unserved customers decide to return when the system re-opens the next morning, there will be a backlogging of demand from day to day. In this case, each day is not independent; starting the simulation with the same initial condition of no demand is not appropriate. The simulation should have each day

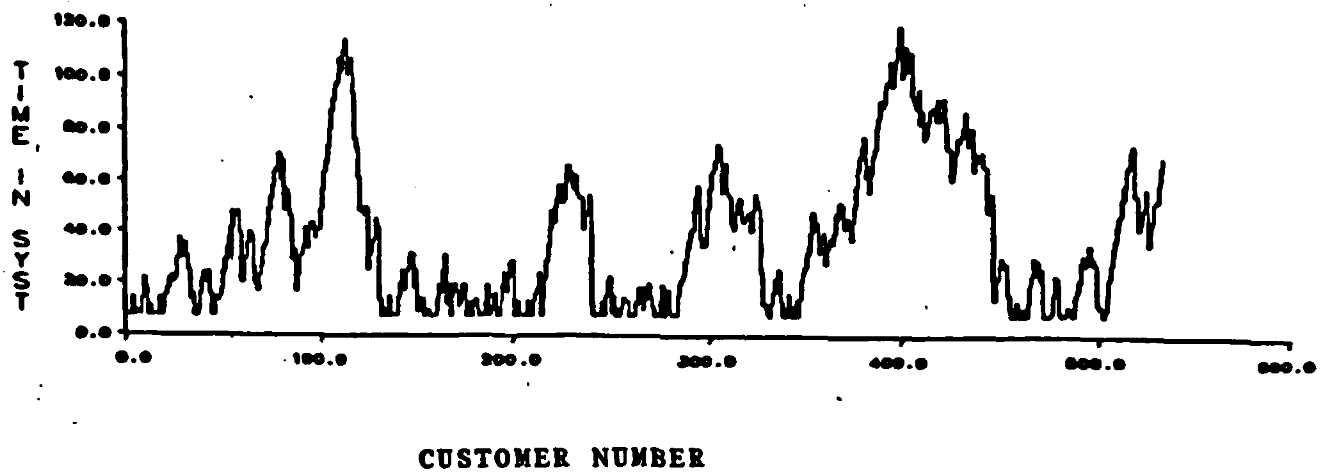


Figure 1 : Time in System for customers in a simple M/D/1 queue.

performances of systems is to retain the current system. This favors the status quo over potential new systems and policies that may in fact be superior.

Unfortunately, many simulation initialization errors are not as transparent as in the above examples. Problems of bias (e.g., underestimated machine utilizations) and spurious secondary correlations in output measures due to correlations with the initial conditions (e.g., congestion correlated with production) can be somewhat difficult to recognize.

Of course, there are situations where initialization effects are not of major concern - for instance, models in which the random variables corresponding to the simulation output are approximately independent and identically distributed. However, in many models, careful examination will indicate that such assumptions necessary to ignore initialization effects are not easily supported.

3. Mathematical effects of initialization

3.1 Estimator bias

We examine the first-order autoregressive process [AR(1)],

$$Y_t = \mu + \alpha(Y_{t-1}-\mu) + \epsilon_t, \quad t = 0,1,2,\ldots,$$

where $|\alpha| < 1$, $\{\epsilon_t\}$ is a sequence of independent identically distributed normal random variables with mean 0 and variance $1-\alpha^2$, and $\mu \equiv \lim_{t\to\infty} E[Y_t]$ is the so-called steady state process mean. [The AR(1) process is a simple model which is useful for illustrating many of the issues in simulation initialization. The lag j serial correlation for this process is $\mathrm{Corr}(Y_i,Y_{i+j}) = \alpha^{|j|}$.]

Suppose that we are interested in estimating μ and that the

initial observation $Y_0 \equiv 0$. Note that $E[Y_t] = \mu(1-\alpha) + \alpha E[Y_{t-1}]$. So an elementary calculation yields $E[Y_t] = \mu(1-\alpha^t)$; the first few Y_t's are quite biased as estimators of μ.

3.2 Estimator variance

The initial conditions also have an effect on the variance of estimators. Consider the familiar M/M/1 queue with traffic intensity less than 1.0. Let $\{Y_i: i=0,1,2,...\}$ denote the customer waiting time process. Suppose again that $Y_0 \equiv 0$. It is noteworthy that the variance of early Y_i's is actually *smaller* than the variance of later Y_i's. I.e., $Var(Y_s|Y_0=0) < Var(Y_t|Y_0=0)$, where s is 'small' and t is 'large.' This phenomenon is illustrated in Figure 2, where we run a number of replications of an M/M/1 queue, each of which is started empty (only a pseudo-random number seed is changed.) Early in each run, the simulated system has not had time to change substantially from its initial empty state.

3.3 Mean squared error of estimator

In point estimation, accuracy (measured by bias) and precision (measured by variance) have been considered important performance criteria. There is usually a trade-off between estimator bias and variance; decreasing one often means that the other will increase. The *mean squared error* (mse) of a point estimator is the squared bias of the estimator plus its variance; the mse is a popular (but more or less arbitrary) criterion for combining the bias and variance into a single quantity [see, for example, Blomqvist (1970), Fishman (1972), Law (1983), and Wilson and Pritsker (1978a)].

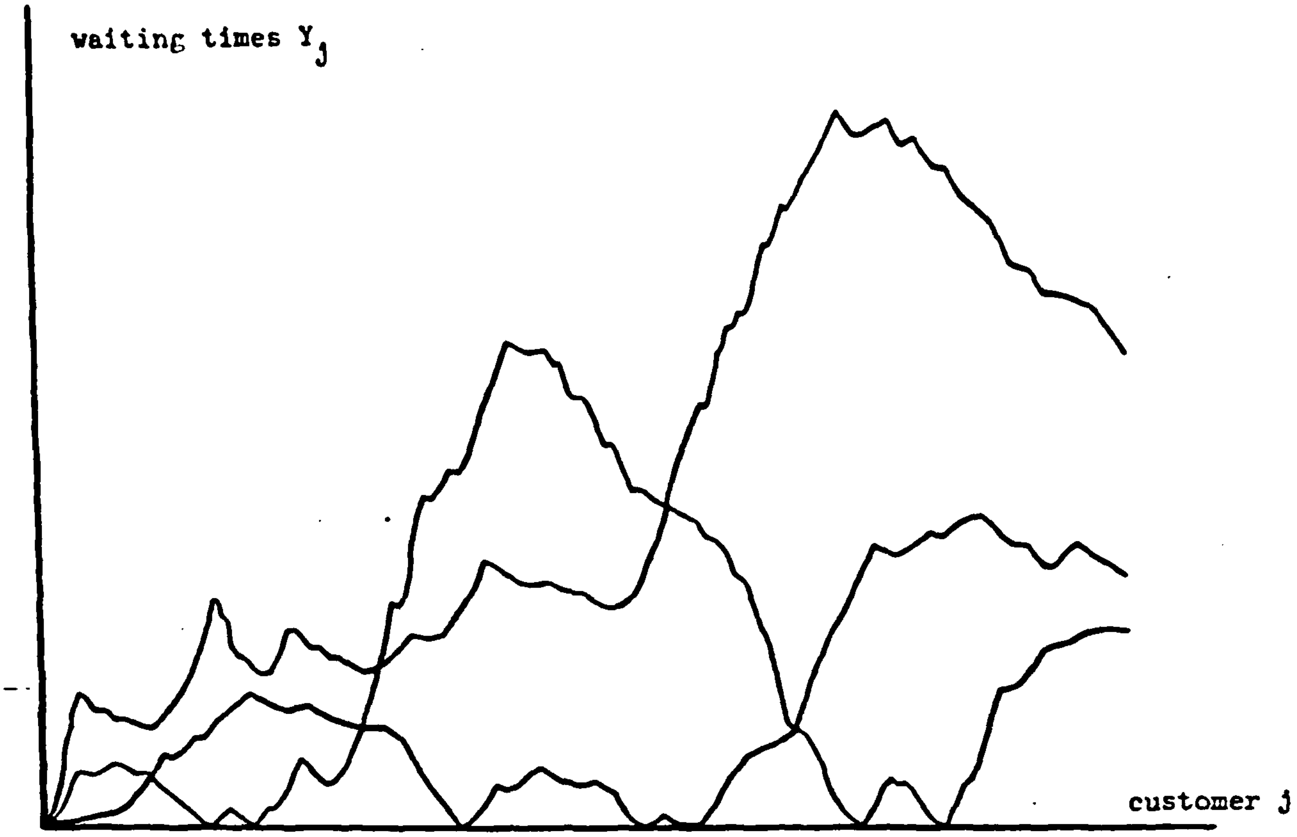


Figure 2: Some realizations of an M/M/1 queue waiting time process. Notice that $V(Y_s | Y_0=0) < V(Y_t | Y_0=0)$ when s is 'small' and t is 'large.'

Fishman (1972) studies the effects of initial conditions on the mse of the random variable corresponding to the sample mean (as an estimator of μ) for the AR(1) process described above. He demonstrates that the common practice of letting the simulation warm up before beginning the collection of data (known as output truncation) is not necessarily advisable (from a mse point of view).

Say that we have two independent output series from an AR(1) process and that one series has been truncated while the other has not been truncated. Snell and Schruben (1979) show that under certain conditions, the non-truncated series yields a mean estimator with a *lower* mse than that from the truncated series. Several points in their paper are worth mentioning. First, consider the space formed by the possible values for α, $Var(\epsilon_t)$, Y_0 (the initial observation), and certain other factors. *If the run is rather long*, there will be a large region of this space where no truncation is called for. However, no matter how long the series is run, there are initial conditions which are so atypical that truncation will reduce the mse. The benefit is reduced as the series becomes longer, but truncation is still of value in some cases. This contradicts the notion that truncation is not beneficial if a simulation is run a very long time [see Madanski (1976)]. It can be less costly to truncate some initial data than to overwhelm poor initial conditions with a long run.

The second point is that, from a mse perspective, positive and negative serial correlations have much the same effect. The region where no truncation is optimal turns out to be larger for positive α than for negative α (because for negative α the sign of the serial correlation alternates with lag). For the same magnitude of α, the amount of positive correlation in the series is greater when this

coefficient is positive than the amount of negative correlation when this coefficient is negative.

Snell and Schruben also derive optimal weightings of mean estimators for the AR(1) process. Least squares estimators (both ordinary and weighted) of the process mean are presented. Data truncation can be viewed as a special case of weighting the output from a simulation. The weight on an observation is zero if it is truncated and one if it is retained for analysis. Expressions are given for finding optimal truncation points for several different estimator performance criteria (in particular, for minimizing the mse.)

3.4 Confidence interval estimators

3.4.1 Batched means and independent replications

Suppose we wish to calculate confidence intervals for the steady state mean $\mu \equiv \lim_{t\to\infty} E[Y_t]$ of some process $\{Y_t\}$. Many experimenters use the method of *batched means*: Divide the process $Y_1,\ldots,Y_n$ into b contiguous batches of m Y_t's each (assume for convenience that m divides n). Hence, $Y_{(j-1)m+1}, Y_{(j-1)m+2},\ldots,Y_{jm}$ constitute the j-th batch, j=1,...,b. Define the batched means:

$$\bar{Y}_j \equiv \frac{1}{m}\sum_{t=1}^{m} Y_{(j-1)m+t}, \quad j = 1,\ldots,b.$$

When applying the method of batched means, the $\bar{Y}_j$'s need to be approximately iid normal random variables with unknown mean and variance. Approximate $100(1-\alpha)\%$ confidence intervals for μ are then given by:

$$\Pr\left\{ \mu \in \bar{\bar{Y}}_n \pm t_{b-1,1-\alpha/2}\, S/\sqrt{b} \right\} \doteq 1-\alpha, \qquad (*)$$

where $\bar{\bar{Y}}_n \equiv \frac{1}{n}\sum_{t=1}^{n} Y_t$ is the random variable corresponding to the grand sample mean, $t_{k,\eta}$ is the upper η quantile of the $t(k)$ distribution, and $S^2 \equiv \sum_{j=1}^{b} (\bar{Y}_j-\bar{\bar{Y}}_n)^2/(b-1)$.

Even if the batched means are, in fact, approximately independent and normally distributed, initialization bias can still cause a number of problems with the above batched means confidence interval estimator. For instance, $\bar{\bar{Y}}_n$ might be quite biased as an estimator of μ; this would result in a confidence interval estimator which is incorrectly centered. Further, S^2 might be a biased estimator for the underlying process variance,

$$\sigma^2 \equiv \lim_{n\to\infty} n\mathrm{Var}(\bar{\bar{Y}}_n).$$

If $E[S^2] >> \sigma^2$, then confidence intervals which are too wide to be of practical use could result. Alternatively, if $E[S^2] << \sigma^2$, the resulting intervals might be invalid [i.e., the probability that they cover μ would be $<< 1-\alpha$].

The method of *replications* may also be unacceptable for use in the presence of initialization bias. This method samples b independent streams (of m observations each) from the stochastic process. The sample means from each of the b replications are calculated, and confidence intervals are formed via (*). Initialization bias is easily seen to be more of a problem here than in the method of batched means - this is because *each* of the b replication streams can be biased by the initial transient (whereas only the single, long stream of the batched means method can have any bias). In view of this argument, Law and Kelton (1982) recommend the use of batched means over independent replications.

3.4.2 Regeneration

The *regenerative* method of confidence interval formulation [cf. Crane and Iglehart (1975) and Crane and Lemoine (1977)] is structured in such a way that it is not directly affected by initialization bias. Regeneration relies on the fact that many (covariance stationary) stochastic processes $\{Y_t, t=0,1,2,\ldots\}$ 'start from scratch' probabilistically at so-called (random) *regeneration times* $0 \leq t_1 < t_2 < \ldots$ The *cycles* $\{Y_t: t_k \leq t < t_{k+1}\}$, $k \geq 1$, are independent and identically distributed.

For $k \geq 1$, define the k-th cycle length $A_k \equiv t_{k+1} - t_k$, and the random variable corresponding to sum of the observations from the k-th cycle,

$$X_k \equiv \sum_{i=t_k}^{t_{k+1}-1} Y_i \ .$$

Clearly, the A_k's and X_k's are iid.

The point estimator for $\mu \equiv \lim_{t\to\infty} Y_t$ is $\hat{\mu} \equiv \frac{1}{n} \sum_{i=1}^{n} Y_i$, where n is the total number of observations taken (starting at time $t_1 \equiv 1$, say). Suppose for convenience that we terminate the simulation after exactly N regeneration cycles; so $n = \sum_{i=1}^{N} A_i$. Then it is easy to see that $\hat{\mu} = \bar{X}_N/\bar{A}_N$, where we use the obvious notation.

Now define $Z_k \equiv X_k - \mu A_k$, $1 \leq k \leq N$; so $E[Z_k] = 0$ and $Var(Z_k) = Var(X_k) + \mu^2 Var(A_k) - 2\mu Cov(X_k,A_k)$. A point estimator for $Var(Z_k)$ is given by:

$\hat{\sigma}_Z^2 \equiv S_X^2 + \mu^2 S_A^2 - 2\mu S_{XA}$, where

$S_X^2 \equiv \sum_{k=1}^N (X_k - \bar{X}_N)^2/(N-1)$, $S_A^2 \equiv \sum_{k=1}^N (A_k - \bar{A}_N)^2/(N-1)$, and

$S_{XA} \equiv \sum_{k=1}^N (X_k - \bar{X}_N)(A_k - \bar{A}_N)/(N-1)$. It can then be shown that approximate $100(1-\alpha)\%$ confidence intervals for μ are given by:

$\Pr\{ \mu \in \hat{\mu} \pm z_{1-\alpha/2}(1/\bar{A}_N)[(S_X^2 + \hat{\mu}^2 S_A^2 - 2\hat{\mu} S_{XA})/N]^{1/2} \} \doteq 1 - \alpha$, where $z_{1-\alpha/2}$ is the upper $1-\alpha/2$ quantile of the Nor(0,1) distribution.

Note that $E[\hat{\mu}] = E[\bar{X}_N/\bar{A}_N] \neq E[X_i]/E[A_i] = \mu$ for *finite* N; Crane and Lemoine give confidence intervals which use *jackknifing* [cf. Miller (1974)] in order reduce the effects of this bias.

Given that the $\{Y_t\}$ process will eventually return to its initial state, the method of regeneration is particularly appealing; in this case, the initial state might serve as a logical choice for regeneration points. However, perhaps the main drawback of this method is that the regeneration cycle lengths might be prohibitively long (especially if the system is initialized in a very atypical state). The reader should peruse, e.g., Bratley, Fox, and Schrage (1983) for a thorough list of advantages and disadvatages of the regeneration method. See Meketon and Heidelberger (1982) for an interesting discussion pertaining in part to initialization effects.

3.4.3 Measures of performance

In the recent simulation literature, confidence interval

estimator performance has been measured using the following criteria [see, e.g., Schmeiser (1982)]:

1) observed interval coverage frequency (of the true population mean, say) for a sample of confidence intervals that use a certain initialization bias control procedure.

2) interval width, typically measured by the sample average half-width of the observed confidence intervals.

3) interval stability, usually measured by the sample variance of the observed interval half-widths.

Some authors also use the coefficient of variation (c.v.) of the interval half-width as another criterion. This can be misleading: the c.v. might be reduced (an apparent improvement) by merely increasing the interval width (not an improvement).

As alluded to previously, initialization bias might result in poor confidence interval performance in terms of all of the above criteria. The effect of bias on interval coverage can be seen using the *coverage function* described in Schruben (1980): Suppose a particular confidence interval estimator is under study, and we obtain a (large) number of independent realizations of that estimator. The *observed confidence level* is defined to be that percent of the realizations which cover the true parameter of interest. The coverage function is simply a plot of the desired confidence level versus the observed confidence level. For instance, Figure 3 [taken from Schruben (1980)] shows how initialization bias might affect the validity of confidence intervals formed using the method of replications - observed coverage levels are generally below the desired coverage levels.

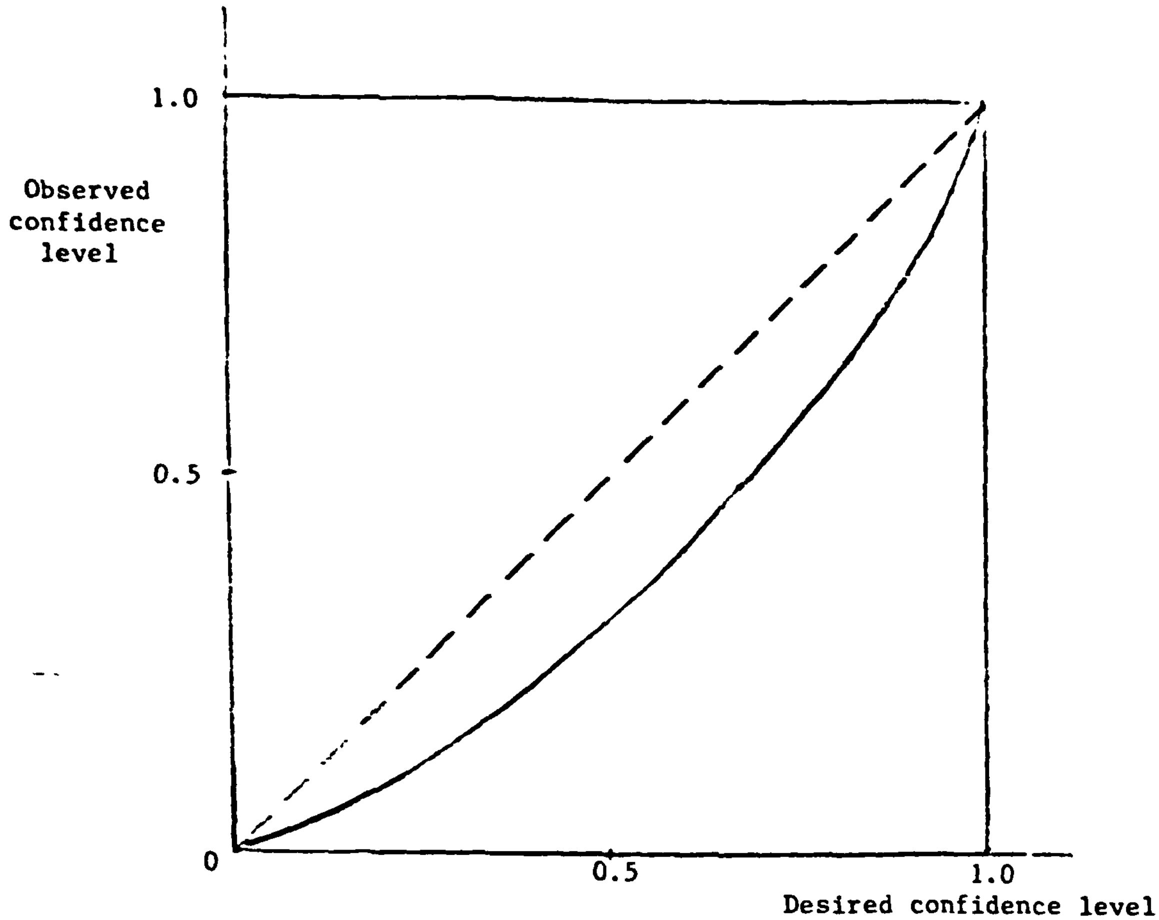


Figure 3: Effect of initialization bias on observed coverage of confidence interval estimator arising from the method of independent replications.

The trade-off between interval half-width and coverage for various confidence interval estimators can be studied using a graphical technique developed in Kang and Schmeiser (1983). They suggest plotting the observed confidence interval center points against the observed half-widths from a sample of confidence interval realizations. The observed coverage frequency is then given by the percentage of such points which fall within 45° of the vertical line drawn at the true population mean μ. In Figure 4a, we plot realizations from a confidence interval procedure inherently yielding low coverage and high interval half-widths; such a plot might result from the presence of initialization bias. Figure 4b illustrates the analogous plot for a superior procedure (i.e., a procedure which has, among other things, mollified the initialization effects); the coverage is close to the desired level and the half-widths are somewhat shorter than those from the previous diagram.

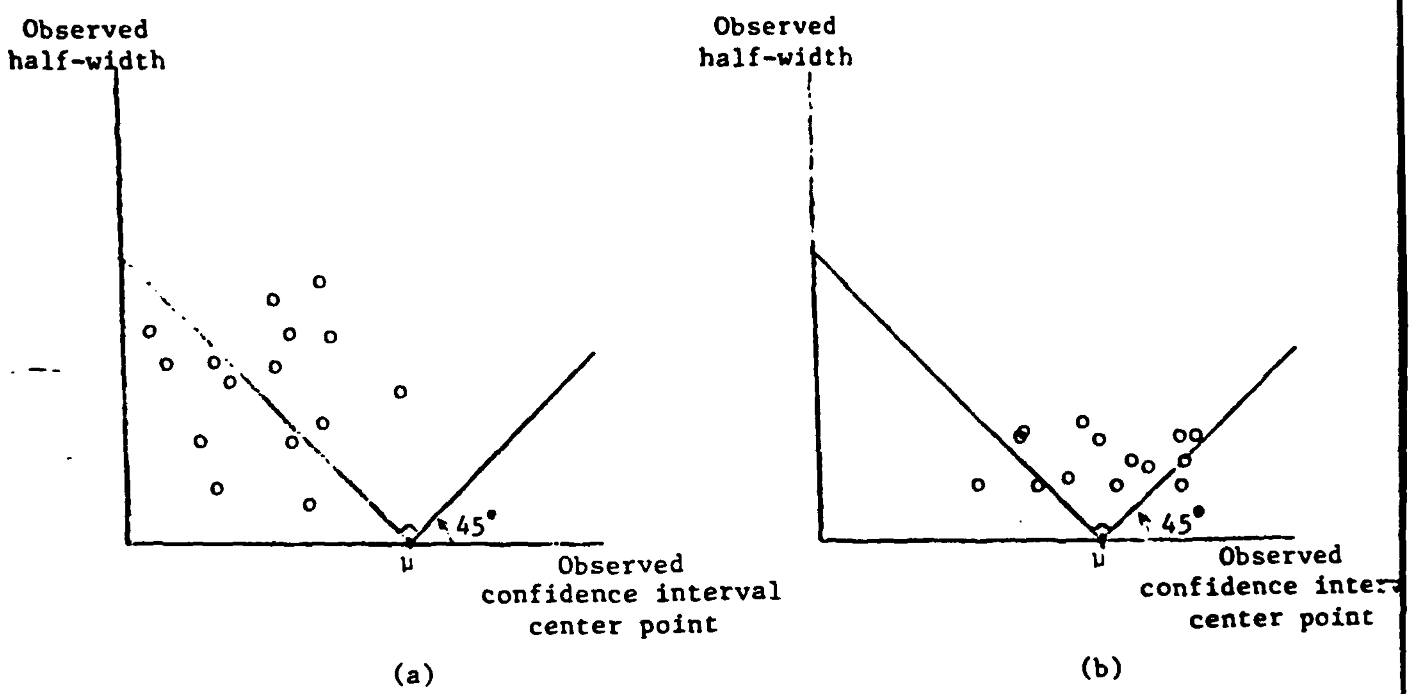


Figure 4: Plots of confidence interval center points vs. half-widths.

(a) A scatter typical of a biased confidence interval estimator (note low values of center points and high half-widths).

(b) A scatter typical of a better confidence interval estimator (note values of center points are close to μ and half-widths are generally lower).

4. Dealing with the problems of initialization

Section 4.1 discusses methods for ascertaining whether or not (significant) initialization bias exists. In Sections 4.2 and 4.3, we give suggestions for dealing with the problems caused by initialization.

4.1 Detection of initialization bias

4.1.1 Graphical methods

Current practice seems to be to simply look at plots of output to try to visually detect any initialization effects. This can be very time consuming in large scale simulation experiments involving many runs, and visually scanning the output might not actually be helpful (recall Figure 1 of this paper). Smoothing the output (by using a moving average, for example) as suggested in Sargent (1979) or Welch (1981) makes visual analysis easier. Averaging observations across several runs (i.e., the first observations from each run are averaged, the second observations are averaged, and so forth) also aids in the detection of initialization effects upon the mean of the output process. The 'CUSUM' type plots in Schruben (1979) are particularly sensitive to changes in the mean of a stochastic process; these plots can be most useful in detecting initialization effects. Transformation of the data (e.g., by taking the logarithms or square roots of the observations) might also help the experimenter in detection of an initial transient.

4.1.2 Tests for initialization bias

Kelton (1980) proposes an intuitively appealing sequential test for bias in which the simulation output regressed on simulated clock time is tested for a zero slope.

Other statistical tests for initialization effects are suggested in Schruben (1982) and Schruben, Singh, and Tierney (1983). In these tests, a quality control viewpoint is taken (to control for inconsistent 'quality' in the output data). Goldsman (1984) gives generalizations of these tests.

Goldsman assumes that the stochastic process $Y_1,\ldots,Y_n$ can be modelled as:

$$Y_i = \mu_i + X_i, \quad i=1,\ldots,n,$$

where $E[Y_i] = \mu_i$, $i=1,\ldots,n$, and the X_i's are stationary. If $\mu_1 = \mu_2 = \ldots = \mu_n$, we say that *no initialization bias is present*; otherwise, bias is present. [The existence of initialization bias in higher order moments is studied in, e.g., Schruben (1981).] The idea behind the tests is to first partition $Y_1,\ldots,Y_n$ into two contiguous, non-overlapping portions. A process variance estimate calculated from the first portion of a realization of $Y_1,\ldots,Y_n$ is then compared to an estimate from the second portion. If it is determined that the two estimates are 'significantly different', the null hypothesis of no bias is rejected.

As an example, we construct some simple tests for bias. Suppose that the first portion of $Y_1,\ldots,Y_n$ is to be divided into b' batches of m observations each, while the second portion is to be divided into $b-b'$ batches. Define:

$$\bar{Y}_{i,j} \equiv \frac{1}{j}\sum_{k=1}^{j} Y_{(i-1)m+k},$$

the average of the first j Y's from batch i, $i=1,\ldots,b$, $j=1,\ldots,m$.

Under mild conditions [see Goldsman (1984)], it can be shown that as m increases,

$$Q_{0,b'} \equiv m \sum_{i=1}^{b'} \left[\bar{Y}_{i,m} - \frac{1}{b'} \sum_{j=1}^{b'} \bar{Y}_{j,m} \right]^2 \xrightarrow{\mathcal{D}} \sigma^2\chi^2(b'-1), \quad 1 < b' < b-1,$$

where σ^2 and $\bar{\bar{Y}}_n$ are defined as in Section 3, $\chi^2(\nu)$ is the χ^2 distribution with ν degrees of freedom, and $\xrightarrow{\mathcal{D}}$ indicates convergence in distribution. Also,

$$Q^*_{0,b-b'} \equiv m \sum_{i=b'+1}^{b} \left[\bar{Y}_{i,m} - \frac{1}{b-b'} \sum_{j=b'+1}^{b} \bar{Y}_{j,m} \right]^2 \xrightarrow{\mathcal{D}} \sigma^2\chi^2(b-b'-1),$$
$$1 < b-b' < b-1.$$

$$Q_{1,b'} \equiv \frac{12}{m^3-m} \sum_{i=1}^{b'} \left[\sum_{j=1}^{m} j(\bar{Y}_{i,m} - \bar{Y}_{i,j}) \right]^2 \xrightarrow{\mathcal{D}} \sigma^2\chi^2(b'), \quad 1 \le b' \le b-1.$$

$$Q^*_{1,b-b'} \equiv Q_{1,b} - Q_{1,b'} \xrightarrow{\mathcal{D}} \sigma^2\chi^2(b-b'), \quad 1 \le b-b' \le b-1.$$

Assuming approximate independence of the batches, we have:

$$F_{0,b'} \equiv \frac{b-b'-1}{b'-1} \frac{Q_{0,b'}}{Q^*_{0,b-b'}} \approx F(b'-1,b-b'-1) \quad \text{and}$$

$$F_{1,b'} \equiv \frac{b-b'}{b'} \frac{Q_{1,b'}}{Q^*_{1,b-b'}} \approx F(b',b-b'),$$

where $F(\nu_1,\nu_2)$ is the F distribution with ν_1 and ν_2 degrees of freedom, respectively.

If we were to use $F_{1,b'}$ to conduct a two-sided test for bias, we would reject the null hypothesis H_0: $\mu_i = \mu$, $\forall i$, at the α level if $F_{1,b'} < f_{b',b-b',\alpha/2}$ or $> f_{b',b-b',1-\alpha/2}$, where $f_{\nu_1,\nu_2,\eta}$ is the upper η quantile of the $F(\nu_1,\nu_2)$ distribution.

Caveats (see the appropriate references for details): Kelton's test has some (minor) problems concerning its applicability.

Goldsman's tests are *asymptotic* and therefore should not be used when the experimenter has only limited data on hand.

4.2 Estimation of steady state parameters

The goal for now is to find an initialization bias control procedure which yields 'good' point estimates and/or confidence intervals for the steady state parameter(s) of interest. If the simulation run is very short, good estimation may not be possible, regardless of the procedure which is used. (Until recently, procedures for dealing with this problem have primarily been heuristic.) In general, one tries to obtain as good an estimator as possible recognizing that without data from the (asymptotic) population of interest, compromise will be necessary.

4.2.1 Truncation rules

There are various approaches to the problem of point/confidence interval estimation. By far the most attention has been given to weighting of the data. Recall that data truncation is a special case of weighting. Truncation allows the simulation to warm up before data are retained for analysis. Many 'truncation rules' have been proposed and studied [see, for example, Gafarian, *et al.* (1978), Kelton (1980), Morisaku (1976), Sargent (1979), and Wilson and Pritsker (1978a,1978b)]; Pritsker and Pegden (1979) cite the following rules (among others):

(a) [Conway (1963)] Let $y_1,\ldots,y_n$ be a realization of the stochastic process under study. Define the *truncation point* t as the number of observations to be deleted from the beginning of the

realization. Choose (the smallest possible) t such that y_{t+1} is neither the minimum nor the maximum of $y_{t+1},\ldots,y_n$.

(b) [Gordon (1969)] Run k independent replications of length n in order to estimate $\mathrm{Var}(\bar{\bar{Y}}_n)$. Choose t equal to the value of n for which the estimate of $\mathrm{Var}(\bar{\bar{Y}}_n)$ begins to drop off as $1/n$.

(c) [Fishman (1973)] Choose t as the smallest n such that the realization $\{y_i\}$ crosses its sample mean $\bar{\bar{y}}_n$ k times (k must be user-specified).

(d) [Schriber (1974)] Consider the k most recent batches of size m:

$$y_{n-im+1}, y_{n-im+2}, \ldots, y_{n-(i-1)m}, \quad i=1,\ldots,k.$$

Define the k corresponding batched means as:

$$\bar{y}(i,n) \equiv \frac{1}{m}\sum_{j=n-im+1}^{n-(i-1)m} y_j, \quad i=1,\ldots,k.$$

Choose t as the smallest n such that all of the $\bar{y}(i,n)$'s are within an interval of length ϵ (k, m, and ϵ must be specified by the user).

(e) [Gafarian, *et al.* (1977)] Choose (the smallest possible) t such that y_t is neither the minimum nor maximum of $y_1,\ldots,y_t$.

The consensus is that simple truncation rules may not in general perform well in all situations. It is not too surprising that such rules do not produce good confidence interval coverage for some of the systems tested in the literature. This observation is particularly

true for many queueing systems (the most popular models for testing simulation procedures). These systems only asymptotically reach steady state and can have persistent serial dependence; none of the data are from the (steady state) population that is being studied. Of interest here would be the lively but inconclusive discussions of Fox (1978), Kleijnen (1979), and Schruben (1978).

4.2.2 Sequential estimation

A straightforward approach for controlling simulation initialization effects is to *sequentially* truncate various amounts of the initial data and see if this changes the decisions indicated by the full output data set. A sequential procedure similar to that suggested in Kelton (1980) might be adapted effectively in this context.

In Heidelberger and Welch (1983), an automatic test for initialization bias based on Schruben (1982) was included in a confidence interval procedure. This resulted in improved performance over an earlier version of the procedure that did not have any initialization error control [see Heidelberger and Welch (1981)]. Such automatic procedures and tests for initialization bias are especially useful when there is a large amount of data to analyze (in which case the analyst does not have the time to plot and look at all of the data). Of course, the condensation of information contained in data (so that it can be understood more easily) is one of the purposes of statistics. The better the information is condensed, the closer the decisions based on sample statistics will be to those based on the entire data set.

4.2.3 Direct modelling

A totally different approach to the estimation problem is to directly model the transient mean function. Snell (1980) uses economic growth models for this purpose. Narasimhan, et *al.* (1982) works with so-called 'intervention time series' models. Richards (1983) uses 'relaxed time series' modelling. More experience is needed with these promising methods before their merits can be judged.

4.3 Experimental design

Another method for dealing with simulation initialization is to treat its effects as a nuisance factor in the design of simulation experiments. For example, the various initial values might be considered as a factor in an ANOVA analysis. Several possible initial states could then be run for each set of experimental factors. One might block such experiments on the initial conditions or regress the output on the initial conditions to try to control this source of error. Consider the beneficial effects in terms of variance reduction that might result from blocking on the initial conditions: Estimators based on the outputs from runs with the same initial conditions could be expected to be positively correlated. Further, estimators from runs with radically different initial conditions might be expected to have negative correlations. See Schruben and Margolin (1978) and Schruben (1979) for a discussion of the relationships among blocking, correlation of the output processes, and variance reduction.

The experimental design approach probably deserves more attention than it has received although it may be of limited practical value in simulation models which have a large number of variables to be initialized.

5. Conclusions

Much progress has been made in dealing with the difficult and important problem of initialization bias. Indeed, simple truncation rules are being replaced by sequential procedures such as those given in Kelton (1980) and Heidelberger and Welch (1983). Various tests to detect the presence of initialization bias are also being developed.

We feel that more attention should be given to the problem in the decision making and experimental design contexts. Simulation researchers might also study the different issues that are involved in simulation experiments with a single system (or single performance measure) vs. experiments with multiple systems (or multiple performance measures). Finally, simulation initialization errors may have different effects depending on whether the simulations are used for design, optimization, evaluation, selection, or feasibility decisions; some investigation in this area is warranted.

References

Blomqvist, N. (1970). On the transient behavior of the GI/G/1 waiting times. *Skandinavisk. Aktuarietidskrift.*, 118-129.

Bratley, P., Fox, B.L., and Schrage, L.E. (1983). *A Guide to Simulation.* Springer-Verlag, New York.

Conway, R.W. (1963). Some tactical problems in digital simulation. *Management Science*, 10, 47-61.

Crane, M.A. and Iglehart, D.L. (1975). Simulating stable stochastic systems: III. regenerative processes and discrete-event simulations. *Operations Research*, 23, 33-45.

Crane, M.A. and Lemoine, A.J. (1977). *An Introduction to the Regenerative Method for Simulation Analysis.* Springer-Verlag, Berlin.

Fishman, G.S. (1972). *Bias considerations in simulation experiments. Operations Research*, 20, 785-790.

Fishman, G.S. (1973). *Concepts and Methods in Discrete Event Digital Simulation.* Wiley, New York.

Fox, B. (1978). Estimation and simulation. *Management Science*, 24, 860-861.

Gafarian, A.V., Ancker, C.J., Jr., and Morisaku, T. (1977). The problem of the initial transient with respect to mean value in digital computer simulation and the evaluation of some proposed solutions. Tech. Report 77-1, Dept. of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA.

Gafarian, A.V., Ancker, C.J., Jr., and Morisaku, T. (1978). Evaluation of commonly used rules for detecting 'steady state' in computer simulation. *Naval Research Logistics Quarterly*, 25, 511-529.

Goldsman, D. (1984). *On Using Standardized Time Series To Analyze Stochastic Processes.* Ph.D. dissertation, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

Heidelberger, P. and Welch, P.D. (1981). A spectral method for confidence interval generation and run length control in simulations. *Comm. Assoc. Comput. Mach.*, 24, 233-245.

Heidelberger, P. and Welch, P.D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31, 1109-1144.

Kang, K. and Schmeiser, B. (1983). Evaluation and comparison of confidence interval procedures. Tech. Memo 83-3, School of Industrial Engineering, Purdue University, West Lafayette, IN.

Kelton, W.D. (1980). The startup problem in discrete-event simulation. Tech. Report No. 80-1, Dept. of Industrial Engineering, University of Wisconsin, Madison, WI.

Kimbler, D.L., Schmidt, J.W., and Davis, R.P. (1979). Bias reduction and initial conditions in digital simulation. Reprint Series No. 7905, Dept. of Industrial Engineering and Operations Research, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Kleijnen, J.P.C. (1979). The simulation startup problem. *Simulation*, 32, 29.

Law, A.M. (1983). Statistical analysis of simulation output data. *Operations Research*, 31, 983-1029.

Law, A.M. and Kelton, W.D. (1982). Confidence intervals for steady-state simulations, I: a survey of fixed sample size procedures. Tech. Report #82-5, Dept. of Management Information Systems, University of Arizona, Tucson, AZ.

Madansky, A. (1976). Optimal initial conditions for a simulation problem. *Operations Research*, 24, 572-577.

Meketon, M.S. and Heidelberger, P. (1982). A renewal theoretic approach to bias reduction in regenerative simulations. *Management Science*, 26, 173-181.

Miller, R.G. (1974). The jackknife - a review. *Biometrika*, 61, 1-15.

Morisaku, T. (1976). *Techniques for Data-Truncation in Digital Computer Simulation.* Ph.D. dissertation, Dept. of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA.

Narasimhan, R., Rubin, P.A., and Melnyk, S.A. (1982). Analyzing simulation output data using time series intervention analysis. Tech. Report, Graduate School of Business, Michigan State University, East Lansing, Mich.

Pritsker, A.A.B. and Pegden, C.D. (1979). *Introduction to Simulation and SLAM.* Systems Publishing Corporation, West Lafayette, Indiana and Halstead Press, a Division of John Wiley & Sons, New York.

Richards, J.E. (1983). *Minimizing Run Length in Simulations Subject to Initial Transients.* Ph.D. dissertation, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA.

Sargent, R.G. (1979). An introduction to statistical analysis of simulation output data. *Proceedings of the AGARD Symposium*, Paris, France.

Schmeiser, B.W. (1982). Batch size effects in the analysis of simulation output. *Operations Research*, 30, 556-568.

Schriber, T. (1974). *Simulation Using GPSS.* Wiley, New York.

Schruben, L.W. (1978). Reply to Fox. *Management Science*, 24, 862.

Schruben, L.W. (1979). Designing correlation induction strategies for simulation experiments. *Current Issues in Computer Simulation.* (Adam, Dogramaci, Eds.) Academic Press, New York.

Schruben, L.W. (1980). A coverage function for interval estimators of simulation response. *Management Science*, 26, 18-27.

Schruben, L.W. (1981). Control of initialization bias in multivariate simulation response. *Comm. of the A.C.M.*, 24, 246-252.

Schruben, L.W. (1982). Detecting initialization bias in simulation output. *Operations Research*, 30, 569-590.

Schruben, L.W. and Margolin, B. (1978). Pseudorandom number assignment in statistically designed simulation and distribution sampling experiments. *JASA*, 73, 504-525.

Schruben, L.W., Singh, H., and Tierney, L. (1983). Optimal tests for initialization bias in simulation output. *Operations Research*, 31, 1167-1178.

Snell, M.K. (1980). *The Application of Regression Methods to the Initial Transient Problem in Computer Simulations.* Ph.D. Dissertation, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

Snell, M.K. and Schruben, L.W. (1979). The use of weighting functions to correct for initialization bias in the AR(1) process. Tech. Report 395, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

Welch, P.D. (1981). On the problem of the initial transient in steady state simulations. I.B.M. Watson Research Center, Yorktown Heights, NY.

Wilson, J.R. and Pritsker, A.A.B. (1978a). A survey of research on the simulation startup problem. *Simulation*, 31, 55-58.

Wilson, J.R. and Pritsker, A.A.B. (1978b). Evaluation of startup policies in simulation experiments. *Simulation*, 31, 79-89.

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION: Unclassified

1b RESTRICTIVE MARKINGS: Unrestricted

2a SECURITY CLASSIFICATION AUTHORITY: Office of Naval Research

2b DECLASSIFICATION/DOWNGRADING SCHEDULE: Not applicable

3 DISTRIBUTION/AVAILABILITY OF REPORT: Unlimited Distribution

4 PERFORMING ORGANIZATION REPORT NUMBER(S): Technical Report No. 594

5 MONITORING ORGANIZATION REPORT NUMBER(S):

6a NAME OF PERFORMING ORGANIZATION: Cornell University

6b OFFICE SYMBOL (If applicable):

7a NAME OF MONITORING ORGANIZATION: Office of Naval Research

6c ADDRESS (City, State and ZIP Code): Ithaca, NY 14853

7b ADDRESS (City, State and ZIP Code): Arlington, VA 22217

8a NAME OF FUNDING/SPONSORING ORGANIZATION: Office of Naval Research

8b OFFICE SYMBOL (If applicable):

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-81-K-0037

8c ADDRESS (City, State and ZIP Code): Arlington, VA 22217

10 SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.

11 TITLE (Include Security Classification): Initialization effects in computer simulation experiments (unclassified)

12 PERSONAL AUTHOR(S): Lee Schruben and David Goldsman

13a TYPE OF REPORT: Interim

13b TIME COVERED: FROM Mar 31 84 TO Aug 25 85

14 DATE OF REPORT (Yr., Mo., Day): 1983 October

15 PAGE COUNT: 25

16 SUPPLEMENTARY NOTATION:

17 COSATI CODES

FIELD	GROUP	SUB GR.

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Issues relating to initializing a simulation program are discussed, and suggested procedures for handling this source of error are reviewed.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21 ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL: Lee W. Schruben

22b TELEPHONE NUMBER (Include Area Code): (607) 256-4856

22c OFFICE SYMBOL: